Bounds of Chain-of-Thought Robustness: Reasoning Steps, Embed Norms, and Beyond

Dingzirui Wang, Xuanliang Zhang, Keyan Xu, Qingfu Zhu, Wanxiang Che, Yang Deng
Harbin Institute of Technology
Singapore Management University

This paper discusses:

How different factors affect the CoT robustness!

What is Your Conclusion?

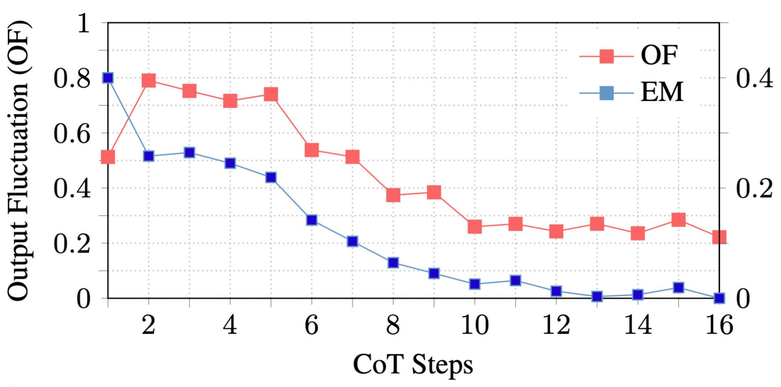
$$\|\varepsilon_K\| \leq \left(A\gamma^K + \frac{C}{1-\gamma}(1-\gamma^K)\right)\|\delta\|$$

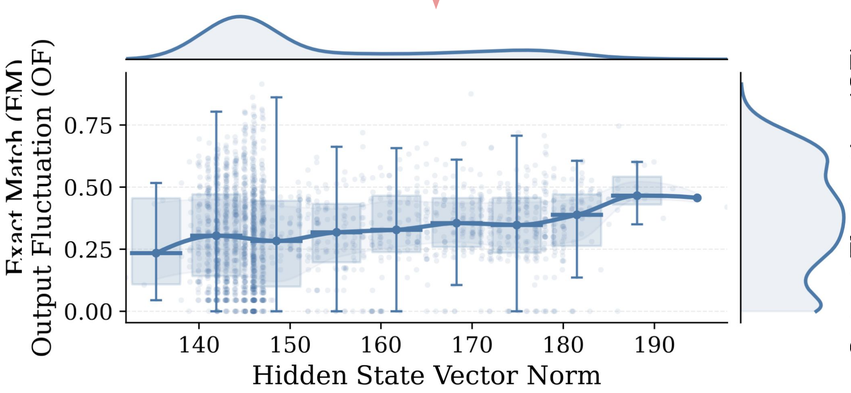
- More CoT **reasoning steps** K can reduce output fluctuation.
- The output fluctuation ε **cannot be entirely eliminated** by increasing the CoT reasoning steps.
- CoT robustness is negatively correlated with the **norms of the input embedding and hidden state vectors** in Transformer.

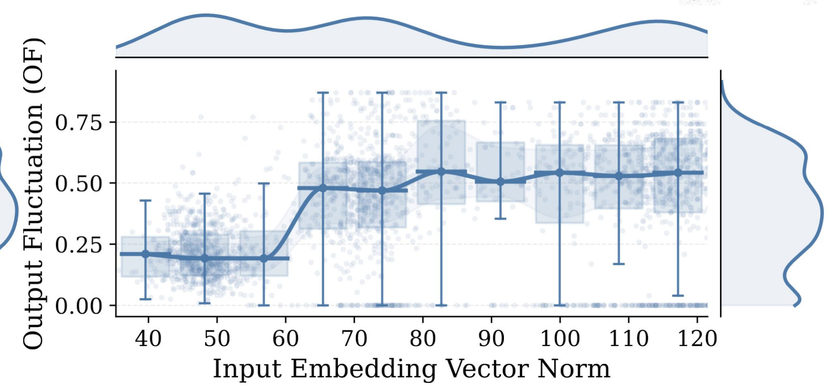
How Experiment Support That?

$$\|\delta\| \leq \frac{(1-\gamma)R}{\eta+\beta}$$

Below are experiments that are consistency with above conclusion.







Can Your Work Guide the Future Design?

Based on our discovery, we propose a prompt optimization method to enhance CoT robutsness, which outperforms baselines with 1.5.

Paper

Code

HomePage